

**UNIVERSITI TEKNOLOGI MARA**

**A STUDY OF INTER-RATER RELIABILITY  
AMONG NOVICE RATERS OF THE BEL120  
WRITING COURSE IN A UNIVERSITY IN  
MALAYSIA**

**NURRUL CHE EME BINTI EMBONG**

**Dissertation submitted in partial fulfillment of the  
requirements for the degree of  
Master of Education in Teaching English as a Second  
Language**

**Faculty of Education**

**May 2013**

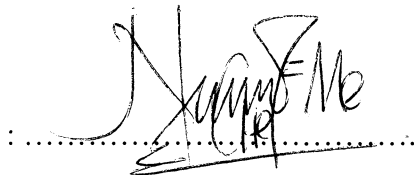
## CANDIDATE'S DECLARATION

I declare that the work in this dissertation was carried out in accordance with the regulation of Universiti Teknologi MARA. It is original and the result of my own work, unless otherwise indicated or acknowledged as referenced work. This topic has not been submitted to any other academic institution or non-academic institution for any other degree or qualification.

In the event that my thesis is to be found to violate the conditions mentioned above, I voluntarily waive the right of conferment of my degree and agree to be subjected to the disciplinary rules and regulations of Universiti Teknologi MARA.

Name of Candidate : Nurrul Che Eme Binti Embong  
Candidate's ID No. : 2009704333  
Programme : M. Ed. Teaching English as a Second Language  
Programme Code : ED720  
Faculty : Education  
Thesis Title : A Study of Inter-Rater Reliability among Novice Raters of the  
BEL120 Writing Course in A University in Malaysia

Signature of Candidate

: 

Date

: 31.05.2013

## ABSTRACT

In the field of writing assessment, various factors have been identified to influence the validity of examinees' scores. One of the most prominent factors believed to threaten scoring validity in an assessment is the raters. Raters played an important role in any type of assessment particularly ones that involve writing. The raters need to be reliable in their rating ability and awarding marks which will be used to determine the examinees ability. Previous research has shown that the most concerned group of raters would be the novice raters. This study seeks to investigate the inter-rater reliability of novice raters in rating using holistic and analytic scoring rubrics for writing assessment, particularly in rating expository essays for BEL120 course in UiTM Dungun, Terengganu. The three novice raters chosen for this study were selected based on the same characteristics. The raters were asked to rate 30 expository essays using Test of Written English (TWE) holistic scoring rubric first and after an interval of two days, they were asked to rate the same essays using BEL120 analytic scoring rubric used by the faculty to mark BEL120 final exam. The marks of the essays were computed and analyzed using SPSS Version 18.0 for Windows to generate the results. Intraclass Correlation Coefficient was used to determine inter-rater reliability since it involved more than two raters. The results showed that the novice raters have low inter-rater reliability level for both scoring rubrics used. The scores discrepancies among the raters also varied greatly thus the raters have low raters agreement when awarding marks using both scoring rubrics. Although the raters were familiar with analytic scoring rubric, the results showed that the inter-rater reliability level for analytic scoring is lower than holistic scoring. The findings provide an insight on the actual level of novice raters' inter-rater reliability and appropriate action is hoped to be taken by Academy of Language Studies in the institution to increase the raters' reliability in rating writing assessment as to increase scoring validity of the examinees.

## ABSTRAK

Dalam bidang penilaian penulisan, pelbagai faktor telah dikenal pasti untuk mempengaruhi kesahihan markah peperiksaan. Salah satu faktor yang paling menonjol yang dianggap boleh mengancam kesahihan pemarkahan dalam penilaian adalah pemeriksa kertas. Pemeriksa kertas memainkan peranan yang penting dalam mana-mana jenis ujian atau peperiksaan terutamanya yang melibatkan penulisan. Pemeriksa kertas haruslah boleh dipercayai dalam keupayaan penilaian mereka dan pemberian markah yang akan digunakan untuk menentukan keupayaan pelajar yang menduduki peperiksaan itu. Kajian sebelum ini menunjukkan bahawa kumpulan yang paling mendapat perhatian dalam kajian berkenaan penilaian penulisan adalah pemeriksa baru. Kajian ini bertujuan untuk menyiasat kebolehpercayaan antara pemeriksa baru menggunakan rubrik holistik dan analitik untuk peperiksaan penulisan, terutamanya dalam karangan bagi kursus BEL120 di UiTM Dungun, Terengganu. Ketiga-tiga pemeriksa atau penilai baru dipilih untuk kajian ini telah dipilih berdasarkan ciri-ciri yang sama. Pemeriksa kertas ini diarahkan untuk menanda 30 karangan menggunakan rubrik Test of Written English (TWE) bagi pemarkahan holistik untuk sesi pertama dan selepas tempoh dua hari, mereka telah diminta untuk menanda karangan yang sama menggunakan rubrik analitik BEL120 yang digunakan oleh fakulti untuk menandakan peperiksaan akhir BEL120. Markah-markah karangan dikira dan dianalisis menggunakan perisian SPSS versi 18.0 untuk Windows bagi menjana keputusan. Intraclass Correlation Coefficient telah digunakan untuk menentukan kebolehpercayaan antara pemeriksa kerana ia melibatkan lebih daripada dua pemeriksa. Hasil kajian menunjukkan bahawa penilai baru mempunyai tahap kebolehpercayaan yang rendah antara penilai untuk kedua-dua rubrik pemarkahan yang digunakan. Markah percanggahan antara penilai juga berbeza-beza dengan itu mempunyai persetujuan yang rendah apabila memberikan markah menggunakan kedua-dua rubrik. Walaupun pemeriksa sudah biasa dengan penggunaan rubrik analitik, keputusan menunjukkan bahawa tahap kebolehpercayaan antara pemeriksa untuk pemarkahan analitik adalah lebih rendah daripada pemarkahan holistik. Hasil kajian memberi gambaran pada tahap sebenar kebolehpercayaan antara pemeriksa-pemeriksa dan tindakan yang sewajarnya adalah diharapkan akan diambil oleh Akademi Pengajian Bahasa di institusi ini untuk meningkatkan kebolehpercayaan pemeriksa baru penilai dalam penilaian penulisan dan meningkatkan kebolehpercayaan markah pelajar yang menduduki peperiksaan.

## ACKNOWLEDGEMENT

First and foremost, my upmost praise to The Almighty Allah S.W.T for giving me the opportunity, strength, courage, health, wisdom and many other things which have helped me to achieve my dream and in completing this dissertation successfully.

My special gratitude and appreciation goes to my supervisor, Dr. Zainab Md. Noor for her supervision and guidance in conducting and completing this dissertation. Her wisdom, patience and dedication during the supervision have helped me in becoming better in doing my research.

I would like to thank Jurainy Jusoh, Nazratul Akmal Hashim and Hazwati Hashim for their cooperation during the research process and for helping in collecting the data. May Allah S.W.T bless you and rewards you for your help.

I would also like to express my gratitude and appreciation to my beloved family members for all the support, patience and understanding that I have received throughout this time. To my beloved husband, Mohd Ridzuan Mat Zin, who have supported me financially, emotionally, morally and for all the hardship that we have to endure, my never ending thank you goes to him. For my beloved little girl, Hannah, who has given me inspiration and courage to keep going further.

My special appreciation goes to my beloved mother, Che Minah Mat Sim for her help throughout this time. I would never be able to do this without her help and no words can be used to describe how grateful I feel for her. My appreciation also goes to my brother, Mohd Ruzed Embong for his support and help and for my sister-in-law, Intan Suhana Idris.

Finally, my special appreciation and gratitude goes to my late father, Embong Ismail for being my inspiration and for giving me the courage to do this even though he is not here with us anymore. May Allah S.W.T bless you with Jannah. Thank you dad. This is dedicated specially for you and mom.